Global identification of noncoding RNAs in *Saccharomyces cerevisiae* by modulating an essential RNA processing pathway

Manoj Pratim Samanta*†‡, Waraporn Tongprasit*‡, Himanshu Sethi*, Chen-Shan Chin§, and Viktor Stolc*¶

*Genome Research Facility, National Aeronautics and Space Administration Ames Research Center, Moffet Field, CA 94035; §Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143; †Systemix Institute, Cupertino, CA 95014; and ‡Eloret Corporation, Sunnyvale, CA 94086

Edited by Sidney Altman, Yale University, New Haven, CT, and approved January 18, 2006 (received for review September 1, 2005)

Noncoding RNAs (ncRNAs) perform essential cellular tasks and play key regulatory roles in all organisms. Although several new ncRNAs in yeast were recently discovered by individual studies, to our knowledge no comprehensive empirical search has been conducted. We demonstrate a powerful and versatile method for global identification of previously undescribed ncRNAs by modulating an essential RNA processing pathway through the depletion of a key ribonucleoprotein enzyme component, and monitoring differential transcriptional activities with genome tiling arrays during the time course of the ribonucleoprotein depletion. The entire *Saccharomyces cerevisiae* genome was scanned during cell growth decay regulated by promoter-mediated depletion of *Rpp1*, an essential and functionally conserved protein component of the RNase P enzyme. In addition to most verified genes and ncRNAs, expression was detected in 98 antisense and intergenic regions, 74 that were further confirmed to contain previously undescribed RNAs. A class of ncRNAs, located antisense to coding regions of verified protein-coding genes, is discussed in this article. One member, *HRA1*, is likely involved in 18S rRNA maturation.

HRA1 | microarray | RNase P | yeast

In *Saccharomyces cerevisiae*, >95% of ribonucleic acids consist of noncoding RNAs (ncRNAs) that perform essential cellular tasks (1). This set includes tRNA and rRNA (which are involved in protein synthesis), small nuclear RNA (snRNA; which performs intron splicing), small nucleolar RNA (snoRNA), RNase P, and RNase MRP (which are active in tRNA and rRNA processing and modification), telomerase RNA (which serves as template during DNA replication), and signal recognition particle (SRP) RNA (which mediates the targeting of proteins to the endoplasmic reticulum). Additional ncRNAs such as microRNA and small interfering RNA (siRNA) were recently discovered in higher organisms, and their roles in regulating developmental pathways through mRNA cleavage are being actively investigated (2).

Recent genome tiling array experiments in *Drosophila*, *Arabidopsis thaliana*, and *Homo sapiens* revealed the widespread presence of short, unannotated transcripts (3–7), some of which could be ncRNAs of unknown cellular function. *S. cerevisiae* is an important model organism for ascertaining their functional roles because most known ncRNA-related pathways in yeast are conserved in higher organisms. Sequence alignment of related subspecies, the primary approach used to identify putative novel intergenic transcripts (8, 9), is unable to locate nonconserved or promoter-based RNA, as illustrated by *SRG1*, a regulatory ncRNA in yeast (10, 11). An alignment-based approach would also fail to identify short transcripts located antisense to protein-coding genes. Therefore, an empirical technique, such as one using genome tiling array technology, is necessary to comprehensively detect short and long transcripts over the entire genome. However, a conventional application of the tiling array approach (3–7) would be unsuccessful in detecting low-abundance transcripts or ncRNAs not transcribed in the chosen cell lines used for array hybridization.

This work circumvented the above problems by combining the strength of tiling array technology with differential gene expression monitoring (12, 13) and observing the changes in global gene expression during modulation of an essential RNA processing pathway (14, 15). High-density tiling microarrays were used to scan the entire yeast genome during cell growth decay that was regulated by promoter-mediated depletion of the *RPP1* gene. Rpp1 is an essential protein shared by both the RNase P and the RNase MRP complexes, and it is functionally conserved from yeast to humans (14). RNase P, an endoribonuclease present in all organisms, removes the 5′ ends of precursor tRNAs to generate mature tRNAs (16, 17). RNase MRP, found only in eukaryotes, is involved in rRNA processing (18, 19). Because RNase P and RNase MRP play fundamental roles in the synthesis of all proteins, disrupting their activities was expected to affect the largest number of cellular pathways and to show widespread differential RNA transcription activities.

Results and Discussion

Nearly 400,000 36-mer oligonucleotide probes, tiling the entire yeast genome including the mitochondrial chromosome with an average gap of 10 bases between two consecutive probes, were synthesized on glass slides by using a maskless array synthesizer (20) (Fig. 1*a*). RNA samples for hybridizing to the arrays were extracted from a conditional lethal allele of *S. cerevisiae* (Table 1), created by placing the *RPP1* gene under control of *GAL10* promoter (14). It allowed the expression of *RPP1* in galactose-containing culture medium but suppressed its expression in glucose-containing medium (Fig. 4, which is published as supporting information on the PNAS web site). A wild-type isogenic strain was used as a control. Both strains were initially grown in galactose-containing medium and subsequently transferred and resuspended into glucose-containing medium. Eight arrays were hybridized with RNA extracted from the Rpp1-depleted cells at 0, 4, 7, 12, 16, 21, and 30 h, and the control cell at 30 h, after initial transfer to glucose-containing medium. The scanned data were normalized by using standard procedures (21). Afterward, 72,633 tiling probes were found to be expressed in absolute or differential sense above a conservatively chosen cutoff. Expressed probes represented genes from 85% of the verified

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviation: ncRNA, noncoding RNA.

Data deposition: The sequences reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database (accession nos. for series: GSE4275; for platform: GPL3471; and for samples: GSM97576, GSM97577, GSM97581, GSM97582, GSM97583, GSM97584, GSM97585, and GSM97586).

¶To whom correspondence should be addressed. E-mail: vstolc@mail.arc.nasa.gov.

© 2006 by The National Academy of Sciences of the USA

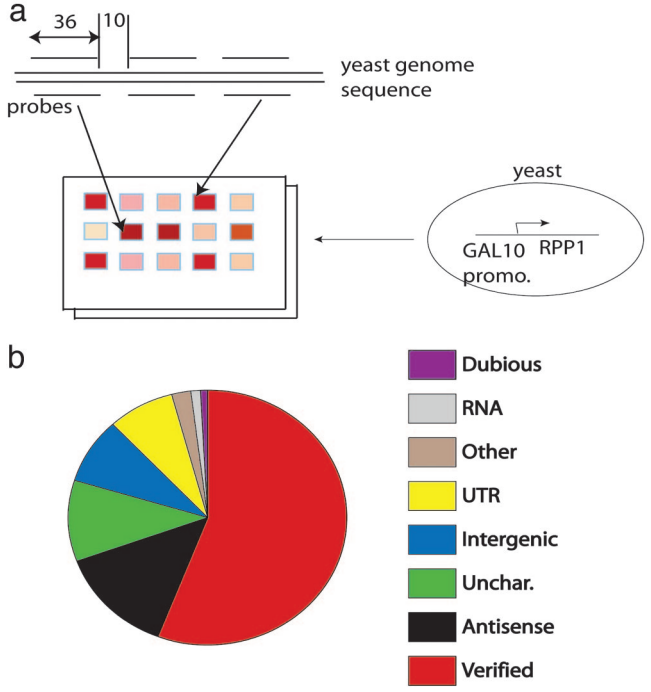

Fig. 1. Experimental design and distribution of the expressed probes. (*a*) A maskless array synthesizer was used to synthesize 36-mer oligonucleotide probes covering both strands of the entire *S. cerevisiae* genome into eight arrays. Arrays were hybridized with RNA from Rpp1-depleted cells. (*b*) The pie chart displays the distribution of 72,633 expressed probes matching different genomic features: red, verified ORF; black, antisense to any annotated feature on the genome; green, uncharacterized ORF; blue, unannotated intergenic region; yellow, untranslated region, including 50-base upstream of ATG sequence and 200-base downstream of stop codon for all verified and uncharacterized ORFs; brown, other genomic features excluding ORF and RNA; gray, nonprotein coding RNA; and violet, dubious ORFs.

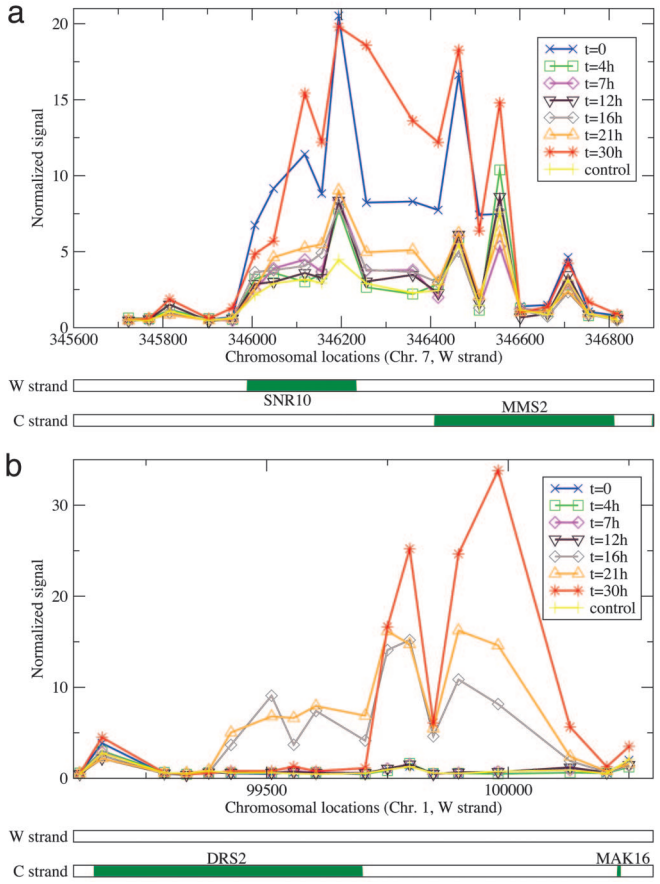

Fig. 2. Transcription of a known and of a previously undescribed ncRNA gene. Hybridization data (*y* axis) are plotted along the genome (*x* axis) displaying signals for ncRNA genes of interest. Lines of different colors represent observations at eight different time points. Although the figure shows data only from the Watson strand, the locations of annotated genes on both strands at nearby regions are shown below each image (green rectangles, annotated genes). (*a*) *SNR10* is a 245-bases-long essential H/ACA small RNA required for modification of rRNA precursor sequence. Based on its mature sequence, transcription at its 3′ end extends almost 400 bases beyond its annotation, suggesting that its precursor RNA is twice as long. The time dependence of the expression pattern for *SNR10* is typical of other ncRNAs. The signal is strong at the first time point (blue), when *RPP1* is overexpressed, but the signal falls immediately (green) to basal level (yellow = control). At later points, the signal increases gradually. (*b*) The previously undescribed RNA *HRA1*, described in this work, is located antisense to ORF *DRS2*. *HRA1* is expressed only at later time points.

protein-coding genes, all tRNA, rRNA, and other known ncRNA genes (22), as well as a large number of antisense and intergenic regions (discussed below).

The gradual depletion of Rpp1 over a defined time course led to inactivation of the RNase P enzyme, thus disrupting the processing of the precursor tRNAs into their mature forms. Because of the short lengths of the tRNAs, typically only one probe measured the total expression for either the precursor or mature tRNA, making it difficult to distinguish between the two. However, the total signal measured by such tRNA-related probes increased over time likely due to rapid accumulation of precursor tRNAs (Fig. 5, which is published as supporting information on the PNAS web site), in accordance with previous studies (14). Moreover, shortage of matured tRNAs was expected to reduce protein synthesis and thus lead to disruption of all cellular processes including transcription. Such transcriptional anomalies were clearly observable at the last time point (t = 30 h; Rpp1-depleted strain) at most transcribed regions of the genome. For protein-coding genes, the signatures of the observed anomaly were (*i*) reduced hybridization signals at the 5′ ends of the genes, possibly by the shortening of the mRNAs due to modulation of *XRN1* expression and (*ii*) increased signals near the 3′ ends, possibly from modulation of exosome activity, by changing the expressions of *RRP43* and *RRP42*. Observed degradation of mRNA could also arise from an absence of translation. No such defect in mRNA quality was seen in the control sample.

All of the known ncRNAs (22–27), encompassing a large spectrum of tRNA, rRNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), RNase P, RNase MRP, and signal recognition particle components, showed strong or time-varying RNA transcription signals (Fig. 6, which is published as supporting information on the PNAS web site). For some ncRNAs (*SCR1* and *SNR10*) and protein-coding genes (*PMP2*), transcription extended well beyond the annotated 3′ ends of their matured forms (Fig. 2*a*). As a characteristic pattern, overexpression of *RPP1* led to increased transcription of known ncRNAs at

Table 1. *S. cerevisiae* strains used in this article

Strain	Genotype
VS164	*MATα GAL1 leu2–3,112 ura3–1 rpp1::LEU2 + pYCpGAL::rppI (URA3)*
VS165	*MATα GAL1 leu2–3,112 ura3–1 rpp1::LEU2 + pYCpGAL (URA3)*

Table 2. Summary information for the previously undescribed transcripts detected by this article

	Total count	Median length, bases	No. of RT-PCR performed	No. of RT-PCR confirmed
All transcripts	98	460	83	74
Antisense transcripts	21	800	8	8
Promoter-based transcripts	50	350	48	43
Intergenic transcripts	27	500	27	23

Depletion of *Rpp1* induced 42 previously undescribed transcripts and repressed 9, and 47 remained unaffected. This information was decided by comparing expression at 16–21 h with expression at 4–12 h.

the first time point ($t = 0$; galactose-containing medium). Within 4 h of transfer to the glucose-containing medium, the same RNA signals adjusted to their basal levels, as determined from the control strain (Fig. 6). The above observation suggests that Rpp1, as a component of RNase P and RNase MRP complexes, not only takes part in the processing of precursor RNA but also affects the overall expression and/or stability of the RNAs. This observation, if confirmed *in vitro* by an RNase P cleavage, could be helpful for determination of putative RNase P substrates.

Apart from the annotated regions of the genome, 21.5% of the expressed probes matched antisense or intergenic segments (Fig. 1*b*), suggesting the possible existence of a large number of putative novel ncRNAs or short ORFs. This result is an unexpected observation because the *S. cerevisiae* genome has been most extensively annotated. Additional screening of the transcribed probes led to the identification of 98 high-confidence ncRNAs (Tables 2 and 3; Figs. 7–18, which are published as supporting information on the PNAS web site), of which 83 were selected for further confirmation by using the RT-PCR method. RNA samples from both $t = 4$ h and $t = 30$ h time points for the Rpp1-depleted cells were used for RT-PCR. By this experiment, 74 previously undescribed RNAs could be confirmed (Table 4, which is published as supporting information on the PNAS web site).

The above list of confirmed previously undescribed ncRNAs includes 21 members located antisense to the protein-coding ORFs. One of them, named *HRA1* (hidden in reading-frame antisense), was located antisense to a previously annotated gene called *DRS2*. Drs2, a Golgi membrane-located transport protein (22), was also determined to be involved in the maturation of 18S rRNA based on a temperature-sensitive mutation-based study (28), but the exact mechanism remained unknown. We suggest below that the characterization of the *drs2* phenotype is likely incorrect, and the observed rRNA processing defect is likely caused by the RNA-coding gene *HRA1* rather than the protein-coding gene *DRS2*.

Drs2 has no known interaction with any protein involved in rRNA synthesis or in the maturation pathway (22). Moreover, none of the four nonessential genes (*BNI1*, *KIP2*, *GIM5*, and *CNB1*) synthetically lethal with *DRS2* is associated with rRNA biogenesis. Instead, all four genes are involved in a common phenotype related to cell membrane trafficking. Because nonessential genes are usually functionally redundant and synthetically lethal with genes of similar function, this observation alone argues against Drs2 being involved in rRNA biogenesis. In contrast, other known ribosomal maturation-related proteins, such as Drs1, have affinity-based interactions or two-hybrid interactions with genes of nucleolar and/or rRNA biogenesis functions.

Reanalysis of the original data used to characterize the *drs2* phenotype, available from ref. 28, and sequencing of the *drs2* locus performed by this work suggest that the observed defect in 18S rRNA maturation may be because of a mutation in *HRA1*, rather than *DRS2*. The earlier study found that a 2.2-kb EcoRI–BglII fragment, including only 528 aa (38% of the coding region) from the 5′ end of *DRS2*, complemented the observed rRNA-related phenotype (Fig. 2*A* in ref. 28). The fragment extended upstream to the 5′ end of *DRS2* and fully included *HRA1* (data not shown). On the other hand, a BamHI–BamHI fragment that included the entire coding region of Drs2, but no additional bases from its 5′ upstream region, did not complement the drs2 phenotype. *HRA1* was only partially included in the second fragment. Moreover, Northern blot analysis of the 2.2-kb EcoRI–BglII fragment DNA, which was successful in complementing the rRNA-related phenotype, detected an additional 1-kb transcript (Fig. 2*B* in ref. 28) similar in length to *HRA1*, based on the array data. The abovementioned Northern blot analysis detected transcripts from both strands of the corresponding genomic DNA. Finally, this work sequenced the 2.2-kb EcoRI–BglII fragment from the mutant *drs2* strain. The sequencing result showed mutations in three consecutive bases that were located within the segment of *DRS2* that contained *HRA1*. The bases 99,563–99,565 of chromosome 1 in yeast were mutated from ATC to TCG.

These results suggest that *HRA1* rather than *DRS2* caused the observed rRNA processing phenotype. These results also highlight the importance of the tiling microarray approach in complementing the classical genetics approach for characterization of genome structure. *HRA1* is the only small nonessential RNA likely to be involved in processing rather than modification of rRNA. Moreover, although small ORFs are known to be encoded within ncRNA-coding genes (e.g., Tar1p in rRNA; ref. 29), the result reported here is a demonstration of the opposite phenomenon: small RNAs being encoded antisense to the protein-coding sequence.

A rare ncRNA-mediated regulatory mechanism in yeast was recently reported by a previous study (10) that found that a promoter-located ncRNA, *SRGI*, regulated the expression of the downstream gene *SER3* through its own transcription. Observations from this study (10) suggest that a similar regulatory mechanism could be more widely present than is currently known. Our data not only confirmed the transcription of *SRG1* (Fig. 3*a*) but also detected 50 additional examples in which an ncRNA was located within the 500-base promoter region of a verified or uncharacterized gene, possibly functioning in the transcriptional regulation of the downstream genes. Of this set, 43 RNAs were positively confirmed by RT-PCR (Table 4) and two examples are shown in Fig. 3*b*. It was found that the

Table 3. Summary information for the previously undescribed transcript distributions among all yeast chromosomes

Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
No. of novel transcripts	3	9	2	16	5	3	9	4	6	5	3	5	8	8	8	4

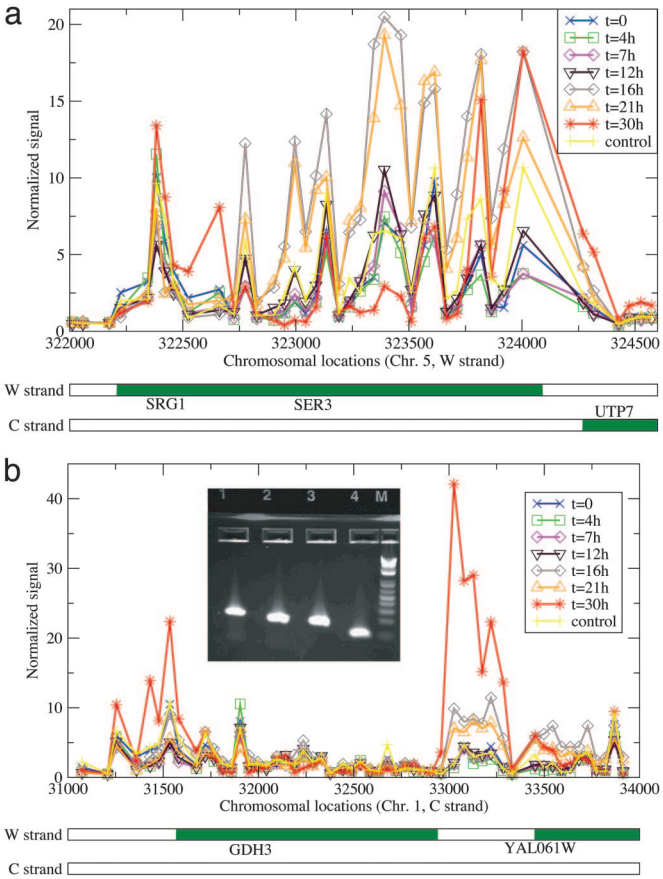

Fig. 3. Promoter-based ncRNA that may function in RNA-mediated transcription regulation. (*a*) *SRG1*, located upstream of *SER3*, regulates the transcription of *SER3* (9). (*b*) Confirmed putative novel RNA *MAN1* and *MAN2*, respectively located upstream of *GDH3* and *YAL061C*, are shown. (*Inset*) Lanes 1 and 2 show RT-PCR results for *MAN1* and *MAN2*. *MAN1* is located between the bases 31,200–31,500 on chromosome 1. *MAN2* is located between the bases 33,000–33,350 on chromosome 1.

expression levels of promoter-based RNA *MAN3* and its immediate downstream gene *PHO5* had a strong inverse correlation (Pearson's correlation coefficient, −0.82; see *Materials and Methods* for details), suggesting that *MAN3* is a negative regulator of *PHO5*.

A phylogenic comparison between four different yeast species was conducted to determine whether the promoter-based transcripts were functionally conserved. DNA sequences for promoter regions of four closely related *Saccharomyces* species were aligned (30), and the conserved regions were ranked as functionally conserved or neutrally conserved by chance based on a computational approach (31). Among the 50 promoter-based transcripts, 21 were functionally conserved between all four yeast species (Z score >3; and Table 5, which is published as supporting information on the PNAS web site). The alignment results, presented in the *Supporting Data Set*, which is published as supporting information on the PNAS web site, can be useful in mutation-based validation studies and for deriving the secondary structures for the corresponding RNAs. We also note that the functional conservation rate of the promoter-based RNAs was not different from other regulatory promoters, confirming our initial premise that the alignment-based computational approaches are unable to identify ncRNAs in the conserved regions of the genome and highlighting the importance of whole genome tiling array measurements.

In summary, this work demonstrates a powerful and versatile method for identifying previously undescribed ncRNAs by modulating an essential RNA processing pathway through depletion of a key ribonucleoprotein enzyme component and monitoring the differential transcriptional activities during the time course of its depletion. The method is applicable to other higher eukaryotes including mammals. In *S. cerevisiae*, this work discovered a class of RNAs located antisense to ORFs. One member, *HRA1*, is likely involved in 18S rRNA maturation. In addition, this work identified and confirmed 73 other previously undescribed ncRNAs, the functions of which, when determined, will provide insights into eukaryotic biology.

Materials and Methods

Design of the Arrays. Eight identical glass-based high-density arrays were constructed by using a maskless array synthesizer (20). Each array contained 388,562 36-mer oligonucleotide probes. Among them, 384,636 tiling probes were chosen from all nuclear and mitochondrial chromosomes of *S. cerevisiae* and 3,926 additional probes were selected to cover the known RNAs at a higher density. Chromosome sequences and annotations used for probe design were downloaded from the *Saccharomyces* genome database (SGD) (21). Tiling probes were selected uniformly from both strands of the chromosomes with average gaps of 10 bases between the consecutive probes on the chromosome. Probes with undesirable features that might have caused difficulties in hybridization were excluded by using an algorithm described in refs. 3–5. All probe sequences and corresponding hybridization data are available from the National Center for Biotechnology Information (NCBI) gene expression omnibus (GEO) database in MIAME format (see data deposition footnote for additional information).

Hybridization Experiment. The arrays were hybridized with total RNA extracted from two *S. cerevisiae* strains. Strain and RNA sample preparation techniques were described in detail in ref. 14. Briefly, a conditional lethal allele was created (VS164, Table 1) by placing *RPP1* under the control of *GAL10* promoter in a plasmid. It allowed expression of *RPP1* in galactose-containing culture medium but suppressed *RPP1* expression in glucose-containing medium. A wild-type isogenic strain (VS165, Table 1), containing pYCp-GAL plasmid, was used as a control. Both strains were initially grown in galactose medium, and subsequently transferred and resuspended into glucose medium. Seven arrays were hybridized with RNA extracted from VS164 strain 0, 4, 7, 12, 16, 21, and 30 h after initial transfer to glucose medium. The remaining array was hybridized with RNA extracted from VS165 30 h after its transfer to glucose medium.

Sample Labeling. Using GIBCO/BRL SuperScript Choice System, total RNA extracted from the yeast cells was converted to double-stranded cDNA. Subsequently, cDNA was labeled by using an oligo(dT) primer containing the T7 RNA polymerase promoter (5′-GGCCAGTAATTGTAATACGACTCACTATAGGGAGGCGG-3′). Briefly, 10 μg of total RNA was incubated with 1× first strand buffer, 10 mM DTT, 500 μM dNTPs, and 5 pM primer for 60 min at room temperature. The second strand was synthesized by incubation with 200 μM dNTPs, 0.07 units per μl DNA ligase, 0.27 units per μl DNA polymerase I, 0.013 units per μl RNase, 1× second strand buffer and 10 units T4 DNA polymerase for 2 h. Double-stranded cDNA was purified by using phenolchloroform extraction and Eppendorf PhaseLock Gel tubes and ethanol precipitated, washed with 80% ethanol, and resuspended in 3 μl water. *In vitro* transcription (IVT) was used to produce biotin-labeled cRNA from the cDNA by using the Ambion (Austin, TX) MEGAscript T7 kit. Briefly, 1 μg double stranded cDNA was incubated with 7.5 mM ATP and GTP, 5.6 mM UTP and CTP, and 1.9 mM bio-11-CTP and

bio-16-UTP (Sigma-Aldrich) in 1× transcription buffer and 1× T7 enzyme mix for 5 h at 37°C. Before hybridization, cRNA was fragmented to an average size of 50–200 bp by incubation in 100 mM potassium acetate, 30 mM magnesium acetate, and 40 mM Tris·acetate for 35 min at 94°C. For quality control at all steps, including input RNA quality, first and second strand cDNA synthesis, *in vitro* transcription, and fragmentation, assay performance was monitored by running small sample aliquots on the Agilent Bioanalyzer (Agilent Technologies, Palo Alto, CA).

Hybridization and Washing. High-density 36-mer tiling arrays were hybridized with 12 μg cRNA in 300 μl, in the presence of 50 mM Mes, 0.5 M NaCl, 10 mM EDTA, and 0.005% (vol/vol) Tween-20 for 16 h at 45°C. Before application, samples were heated to 95°C for 5 min, then 45°C for 5 min, and then centrifuged at 12,000 × *g* for 5 min. Hybridization was performed in a hybridization oven with continuous mixing. After hybridization, arrays were washed in nonstringent (NS) buffer (6× saline-sodium phosphate-EDTA; 0.01% Tween-20) for 5 min at room temperature, followed by washing in stringent buffer (100 mM Mes, 0.01 M NaCl, and 0.01% Tween-20) for 30 min at 45°C. After washing, arrays were stained with streptavidin-Cy3 conjugate (Amersham Pharmacia) for 25 min at room temperature, followed by a 5-min wash in NS buffer, a 30-s rinse in final rinse buffer, and a blow-dry step using high-pressure grade 5 argon.

Scanning and Normalization. Arrays were scanned on an Axon 4000B scanner and features were extracted by using NIMBLESCAN software (NimbleGen Systems, Madison, WI). A quantile normalization procedure was applied to normalize data from different arrays (22). In this method, probes from all arrays were mapped to a reference distribution, which itself was created by taking averages of the sorted raw data (log base 2) from all eight arrays and scaled to a median of one. A statistical analysis of genome-wide data sets was performed by plotting joint distributions of log-scaled raw data from each pair of data set charts (available on request). Also, pair-wise correlation coefficients between the data sets were computed. All pairs showed strong correlations (Pearson coefficient, 0.82–0.97), except for those sets containing data measured 30 h after depletion of Rpp1 was initiated (SET7). Different normalization schemes (e.g., quantile normalization on only seven unaffected sets) were tested with SET7 to account for this unusual distribution, however, none resulted in significant differences.

GC Content of the Probes. Normalized probe signals were further adjusted to take into account biases due to probe GC contents. Median signals were computed for probes located within verified genes and with identical numbers of GC nucleotides. Signals for probes with higher GC contents were stronger on average. Log-normalized data of all probes were adjusted with a corrective factor, which took such GC variation into account.

Reference Set of Unexpressed Probes. A reference set of 14,830 probes was created by considering all promoter regions of verified genes that did not have overlaps with other annotated ORFs, RNA, or repeat regions. Promoter regions were considered to be 50–300 bases upstream from the 5′ ends of the verified ORFs. Before determining overlaps, all annotated features were extended by 50 bases at both ends to account for precursor transcripts being longer than the matured forms.

Filtering of the Expressed Genes. Two different procedures were used to derive probes with significant activities. They were both based on the null set of 14,830 probes derived from the promoters of verified genes.

(i) Filtering based on absolute signal intensity. A cutoff was chosen so that only 2.5% of the probes selected from the promoter regions had absolute signal above it. All probes that had at least one measurement above the cutoff were chosen as positive.

(ii) Filtering based on differential activity. The difference between the maximum and minimum signal from all time-points were computed and a cutoff was chosen following the same criteria as above. All probes that had variations greater than the cutoff were included. The same procedure was followed for the differences between the second highest and the second lowest numbers. In total, 61,020 probes were selected based on absolute expression, and only 11,613 additional probes were chosen solely based on differential activities.

Cross-Hybridization. To check whether some of the observed activities in the intergenic regions were because of cross-hybridization from the mismatched probes, two methods were used. The first method determined which probes aligned with multiple regions of the genome when differences of up to two bases were considered. The second method searched for smaller overlaps between segments of the probes and the entire genome by using a "frequency parameter" computed for all probes. The details on computation of frequency parameter have been discussed in refs. 3–5.

High-Confidence Transcripts. A subset (37,872) of probes was selected from the group of 72,633 expressed probes based on a more stringent cutoff that allowed inclusion of only 1% of the promoter-based probes. The list was further screened to discard any probes matching annotated or repeat regions. The remaining probes were combined into 1,391 longer transcribed regions and were manually screened based on the lengths of transcripts, signal intensities, and differential activities. A high-confidence list of 98 transcripts was created (Table 4 and Figs. 7–18).

RT-PCR. Eighty-three candidates were chosen from the high-confidence list for further verification via RT-PCR. Oligonucleotide primer pairs (20- to 25-mers) were designed from the RNA sample by using the Massachusetts Institute of Technology Primer3 online server (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and are listed in Table 4. RNA samples from two time points (t = 4 h and t = 30 h; Rpp1-depleted cells) were converted to ssDNA, and the primers were used to amplify cDNA by using reverse transcriptase. Positive identification by RT-PCR meant that the largest ethidium bromide-stained band observed in an agarose gel corresponded to the correct size for the transcript.

Correlation Analysis for MAN3. To decide whether the expression of *MAN3* and the downstream gene *PHO5* (Fig. 7) were inversely correlated, the Pearson correlation coefficient was computed for the log-normalized average expression levels. For *MAN3,* the average was computed for the probes located between 431,350 and 431,650 on the Watson strand of chromosome 2. For *PHO5*, only the probes near its 3′ end, between bases 429,700 and 429,900 on the Crick strand of chromosome 2, were considered because those probes best represented the differential expression of the gene. The anomalous SET7 (t = 30 h) was excluded from the calculation.

We thank John L. Woolford, Jr., for the *drs2* yeast strain and P. McCue from the Universities Space Research Association (Houston) for critical reading of the manuscript. This work was supported by grants from the National Aeronautics and Space Administration (NASA) Center for Nanotechnology; the NASA Fundamental Biology Program; the Computing, Information, and Communications Technology programs (Contract NAS2-99092) (all to V.S.); Soli Deo Gloria.

1. Peng, W. T., Robinson, M. D., Mnaimneh, S., Krogan, N. J., Cagney, G., Morris, Q., Davierwala, A. P., Grigull, J., Yang, X., Zhang, W, *et al.* (2004) *Cell* **113,** 919–933.
2. Bartel, D. P. (2004) *Cell* **116,** 281–297.
3. Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M. F., Rifkin, S. A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P. E., *et al.* (2004) *Science* **306,** 655–660.
4. Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., Rinn, J. L., Tongprasit, W., Samanta, M., Weissman, S., *et al.* (2004) *Science* **306,** 2242–2246.
5. Stolc, V., Samanta, M. P., Tongprasit, W., Sethi, H., Liang, S., Nelson, D. C., Hegeman, A., Nelson, C., Rancour, D., Bednarek, S., *et al.* (2005) *Proc. Natl. Acad. Sci. USA* **102,** 4453–4458.
6. Shoemaker, D. D., Schadt, E. E., Armour, C. D., He, Y. D., Garrett-Engele, P., McDonagh, P. D., Loerch, P. M., Leonardson, A., Lum, P. Y. & Cavet, G. (2001) *Nature* **409,** 922–927.
7. Yamada, K, Lim, J., Dale, J. M., Chen, H., Shinn, P., Palm, C. J., Southwick, A. M., Wu, H. C., Kim, C., Nguyen, M., *et al.* (2003) *Science* **302,** 842–846.
8. McCutcheon, J. P. & Eddy, S. R. (2004) *Nucleic Acids Res.* **31,** 4119–4128.
9. Tupy, J. L., Bailey, A. M., Dailey, G., Evans-Holm, M., Siebel, C. W., Misra, S., Celniker, S. E. & Rubin, G. M. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 5495–5500.
10. Martens, J. A., Laprade, L. & Winston, F. (2004) *Nature* **429,** 571–574.
11. Inada, M. & Guthrie, C. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 434–439.
12. Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. & Futcher, B. (1998) *Mol. Biol. Cell* **9,** 3273–3297.
13. DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997) *Science* **278,** 680–686.
14. Stolc, V. & Altman, S. (1997) *Genes Dev.* **11,** 2926–2937.
15. Mnaimneh, S., Davierwala, A. P., Haynes, J., Moffat, J., Peng, W. T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A., *et al.* (2004) *Cell* **118,** 31–44.
16. Gopalan, V., Vioque, A. & Altman, S. (2002) *J. Biol. Chem.* **277,** 6759–6762.
17. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. (1983) *Cell* **35,** 849–857.
18. Schmitt, M. E. & Clayton, D. A. (1993) *Mol. Cell. Biol.* **13,** 7935–7941.
19. Chu, S., Archer, R. H., Zengel, J. M. & Lindahl, L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 659–663.
20. Nuwaysir, E. F., Huang, W., Albert, T. J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J. P., Ballin, J., *et al.* (2001) *Genome Res.* **12,** 1749–1755.
21. Bolstad, B. M., Irizarry R. A., Astrand, M. & Speed, T. P. (2003) *Bioinformatics* **19,** 185–193.
22. Christie, K. R., Weng, S., Balakrishnan, R., Costanzo, M. C., Dolinski, K., Dwight, S. S., Engel, S. R., Feierbach, B., Fisk, D. G., Hirschman, J., *et al.* (2004) *Nucleic Acids Res.* **32,** D311–D314.
23. Li, B., Nierras, C. R. & Warner, J. R. (1999) *Mol. Cell. Biol.* **19,** 5393–5404.
24. Lygerou, Z., Mitchell, P., Petfalski, E., Seraphin, B. & Tollervey, D. (1994) *Genes Dev.* **8,** 1423–1433.
25. Venema, J. & Tollervey, D. (1999) *Annu. Rev. Genet.* **33,** 261–311.
26. Chamberlain, J. R., Kindelberger, D. W. & Engelke, D. R. (1996) *Nucleic Acids Res.* **24,** 3158–3166.
27. Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R. H. & Engelke, D. R. (1998) *Genes Dev.* **12,** 2463–2468.
28. Ripmaster, T. P., Vaughn, G. P. & Woolford, J. L., Jr. (1993) *Mol. Cell. Biol.* **12,** 7901–7912.
29. Coelho, P. S., Bryan, A. C., Kumar, A., Shadel, G. S. & Snyder, M. (2002) *Genes Dev.* **16,** 2755–2760.
30. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. (2003) *Nature* **423,** 241–254.
31. Chin, C.-S., Chuang, J. H. & Li, H. (2005) *Genome Res.* **15,** 205–213.